I Know You by Heart: Biometric Authentication based on Electrocardiogram (ECG) signals

Christoph Lipps[1], Lea Bergkemper[1], Jan Herbst[1] and Hans Dieter Schotten[1,2]
[1]German Research Center for Artificial Intelligence, Intelligent Network Research Group, Kaiserslautern, Germany
[2]University of Kaiserslautern, Division of Wireless Communication and Radio Positioning, Kaiserslautern, Germany
Christoph.Lipps@dfki.de
Lea.Bergkemper@dfki.de
Jan.Herbst@dfki.de
Hans_Dieter.Schotten@dfki.de

Abstract: Trust, confidence and trustworthiness are of fundamental importance in human societies, usually established and sustained through personal relationships. But, due to the globalization and ongoing interconnection of everything up to Cyber-Physical Productions Systems (CPPS) and the Internet of Everything (IoE), physical attendance is no longer necessary. (Remote) access to systems is possible from anywhere on the globe. Accompanied with the lack of personal relationship is the challenge to trust entities –humans or machines-, and proof the identity they are claiming to be. Whether it's payment transactions with smartwatches, logging in to systems, or accessing sensitive parts of buildings, the user's identity is the basic prerequisite. For human participants, for instance, the verification can be obtained through biometrics. These are distinguishable into physiological, biological and behavioral features, each characteristic but of varying difficulty to deduce them. Although using biometric features is not a new concept -indeed they are the oldest form of authentication-, modern approaches are shifting them back into focus. Improved sensor technology enables the identification of people by their gait, or to distinguish them by their characteristic gestures. This work highlights how the availability of (medical) data, and the possibilities of Artificial Intelligence (AI) contribute to the identification and authentication of humans. Therefore, Electrocardiogram (ECG) signals are recorded using a Microcontroller Unit (MCU) and ECG electrodes to derive a three-lead ECG. Using different Machine Learning (ML) algorithms: K-Nearest Neighbor (KNN), Support Vector Machines (SVM) and Gaussian Naïve Bayes (GNB); it is analyzed whether the ECG signals are able to distinguish individuals. Thereby, the ML algorithms are compared with each other, determining which one achieves the best results. The results of the evaluation indicate that ECG signals are capable to distinguish humans based on their heartbeat in such a manner that they can be used as Human - Physically Unclonable Functions (Human-PUFs). Furthermore, the results give reason to assume that the algorithms can also be used for medical applications, for example to recognize heart diseases.

Keywords: Biometric Authentication; Human-PUFs; Physically Unclonable Functions; Physical Layer Security; Industrial Internet of Things; Trust

1. The Industrial Internet of Things and the matter of Authentication

The proliferation of broadband communications and the accompanying inter-connectivity: the coupling of everything with everything in Internet of Everything (IoE) and Industrial Internet of Things (IIoT) scenarios; is accelerated by recent developments in Artificial Intelligence (AI), the availability of large amounts of data and the ability to acquire, analyse and draw conclusions from this data. In combination with cloud computing, Cyber-Physical Production Systems (CPPS) and the virtualization of processes and systems, towards Digital Twins (DT) (Barricielli, et al., 2019), the analogue world is gradually being mapped into their virtual representative. In the process, however, the technological boundaries, places of operation and frontiers are blurring as well. There is no need of physical proximity to access a system, it can be accessed from anywhere in the world.

But this is also accompanied by strict requirements for the (cyber)security of these systems. The various scenarios –remote, on-site-, and entities involved –humans, machines, services-, though, pose major challenges. Although the fundamental requirements of any security concept are the same: the confidentiality integrity and availability (Schneier, 2015), the implementation is basically different. There is a difference between authenticating a human moving between several buildings and validating a machine statically located at one position in a factory or an Automated Guided Vehicle (AGV) moving independently in this factory. **Figure 1** shows this interactivity and the meshing of various components and demonstrates the difficulty of finding a holistic solution for these requirements.

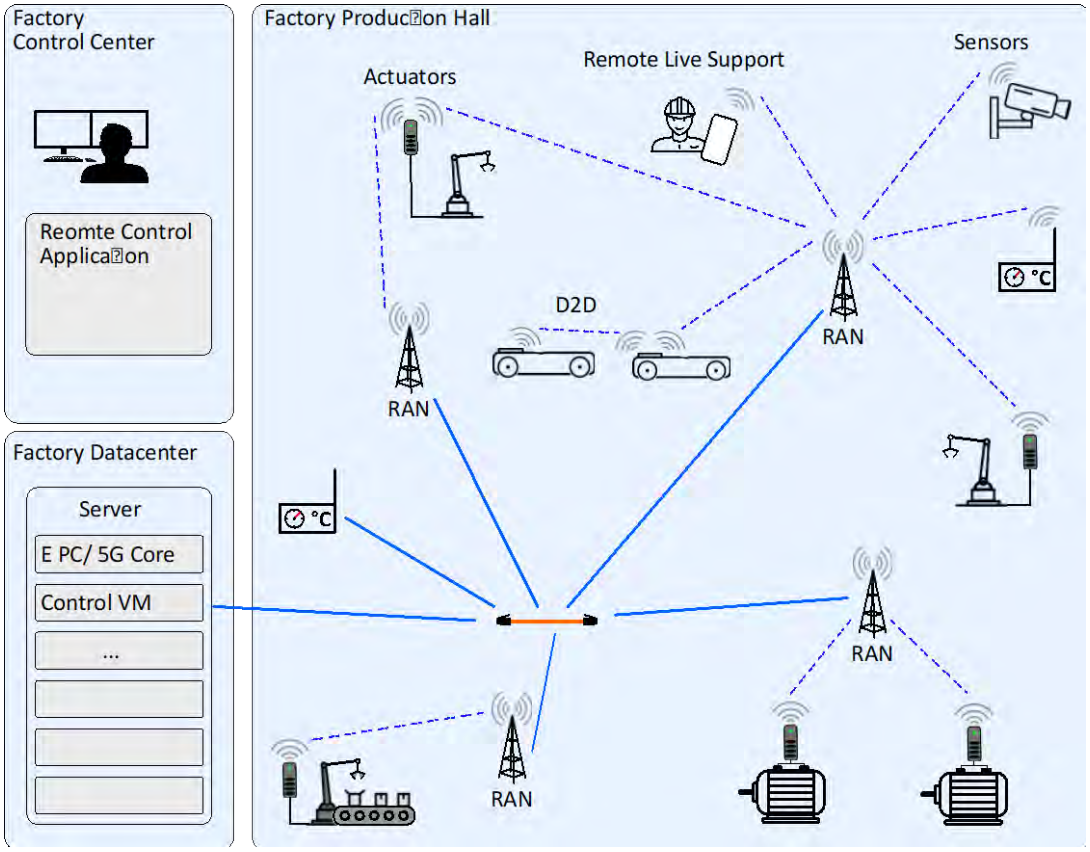

Figure 1: Industrial Internet of Things environment with wirelessly connected entities, including M2M, M2S and H2M applications

When it comes to Machine-to-Machine (M2M) and Machine-to-Service(M2S) applications, the approaches are focused on certificates, Hardware Security Modules (HSMs) and Trusted Platform Modules (TPMs), as well as encryption in general. Besides there are approaches with respect to Physical Layer Security (PhySec) aiming to utilize various physical characteristics of semiconductors (Lipps, et al., 2018) and the wireless propagation channel (Lipps, et al., 2020b) to authenticate devices and to derive cryptographic primitives. For all human-related scenarios, by contrast, biometric methods are (actually exclusively) appropriate. As the operating principle -the utilization of inherent/intrinsic characteristics- is analogous to PhySec, it is referred to as Human - Physically Unclonable Functions (Human-PUFs) (Lipps, Bergkemper & Schotten, 2021a).

The remainder of this work is organized as follows. In Section 2 an overall introduction to the authentication factors and their relevance in Multi-Factor Authentication (MFA) scenarios is given. Section 3 gives a brief insight into Artificial Intelligence and the global impact it enables, whereas Section 4 describes the setup for measuring the ECG values and the performance of the measurement sequences. A classification and discussion of the results is provided in Section 5. Section 6 concludes the work and provides an outlook on future work and the next steps to the further enhancements.

2. Biometric Authentication: An Evolution

The utilization of biometric features is not merely one of the simplest forms of authentication, it is actually the oldest form of identification. For instance, humans recognize each other by smell, voice, behaviour, stature, and facial features, even across distances and in groups, although only some or a few of the features are available. We recognize each other on the phone, although only phonetic and morphological features are present, which are additionally technically compressed and transmitted via (bad) channels. This uniqueness of biometric features is long known and have been used in forensics for more than 100 years, for instance in the context of fingerprints (Schneier, 2014).

But, increasing globalization and interconnectivity of environments and applications render this identification and authentication more challenging than ever before. Involved are no longer just humans in traditional Human-to-Human (H2H) scenarios, but also Human-to-Machine (H2M), Machine-to-Service (M2S), and all other forms

of mutual interaction. Furthermore, distances are also becoming wider, entities are no longer necessarily in close proximity and in the same area, remote access to systems is possible from anywhere on the globe. Altogether, the art of authentication changed, it grew more complex, while at the same time the technical possibilities increased significantly. In terms of biometric authentication, this implies that algorithms have significantly more difficulty in assigning human characteristics and detecting attempts of deception. However, modern approaches for biometric authentication are shifting them back into focus: methods of Artificial Intelligence (AI) enable the analysis of data and the ability to draw relevant conclusions from it; and improved sensor technology enables the identification of people by their gait, distinguish them by their characteristic gestures and by their heart rate.

In order to consider a technical solution of biometric authentication, it is reasonable to first understand the meaning of the term, which is why a short explanation is provided. Following the National Institute of Standards and Technology (NIST) authentication is "a process of establishing confidence" (Burr, et al., 2013). Combined with the proposal of *O'Gorman* as a "process of positively verifying the identity" of an entity (O'Gorman, 2003), an adequate indication can be made. "Establishing confidence" and "verifying the identity" form the basis of a trustworthy communication and exchange of information.

2.1 Authentication Factors: About Knowledge, Inherence and Possession

For the authentication of entities there are, as already mentioned, various methods, depending upon what should be authenticated in detail. On a first level, the methods are divided into three areas, as indicated on the left-hand side of **Figure 2**:

- Knowledge,
- Inherence, and
- Possession.

The *knowledge-based* factors encompass *things that someone knows*, such as passwords, Personal Identification Number (PIN) codes or secret patterns, whereas *inherence* factors describe *something that someone is*. These include factors based on physiological or biological characteristics. Factors *owned* by someone are referred to as *possession factors.* These are typically physical things such as key cards, token and metal keys.

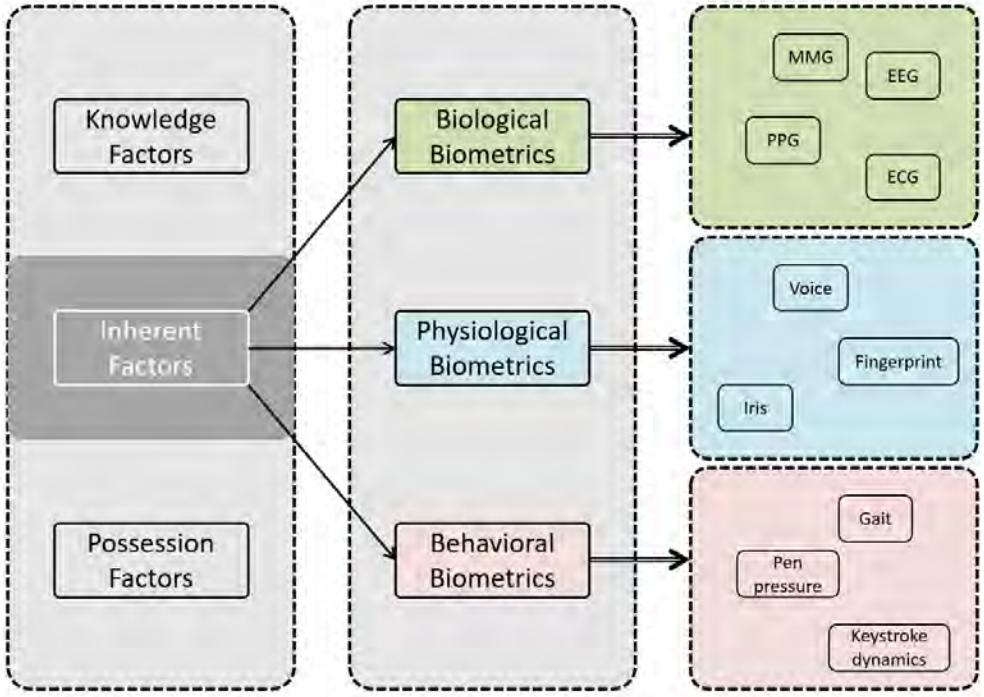

Figure 2: Authentication Factors are basically divided into three levels: Knowledge-, Inherence- and Possession Factors. Inherence Factors are further specified into the sub-groups: biological-, physiological- and behavioural biometrics, which could be further specified

Besides the mentioned categories there are approaches considering additional properties summarized by *context information.* This is additional information such as the geographical location, the Internet Protocol (IP) address and the status on an entity (Duque Anton, et al., 2018).

2.2 The Combination of Factors: Multi-Factor Authentication

As mentioned before, there are different groups of factors, which are differently suitable (or not) for individual applications. Moreover, there are applications with differently stringent security requirements. In its simplest form, this is just a single factor, such as a password or PIN code. This is called Single-Factor Authentication (SFA). With increasing demands on confidentiality and value of an access, the level of security can be increased by coupling such a password/PIN with a second factor to form a Teo-Factor Authentication (2FA). Commonly, a factor from another group of factors is used for this purpose: if the first factor is *Knowledge*-based, for instance, the second factor should be either *Possession*-based or an *Inherence*-factors. As an example, this is the standard for withdrawing money at an Automated Teller Machine (ATM). The combination of bank/credit card (possession) and PIN code (knowledge) is required.

For a Multi-Factor Authentication (MFA), another additional factor is combined to obtain a *strong authentication*. Factors from each of the groups must be combined and validated together (Ometov, et al., 2018). To stick with the ATM example, withdrawing money could be made more secure by integrating an additional biometric factor into the authentication process. The possession of the card and the knowledge of the PIN is not a reliable validation of the identity of a person, it simply indicates that a person knows the PIN code associated with the card.

2.3 Human Characteristics: Biometrics

As discussed in Section 2.2, current biometric methods offer additional benefits in terms of further enhancing the level of security. Nevertheless, when using them it is important to keep in mind that they are “powerful and useful, but they are not keys” (Schneier, 1999). Unlike mechanical keys, the human body is a living organism, which is constantly changing: cells are dying and new ones are being formed (Renaud, 2005). The resulting deviations and inaccuracies of the parameters must be considered in a technical validation.

As indicated in **Figure 2**, biometrics are attributed to the "inherence" factors. As illustrated in the central block of the figure, these can in turn be differentiated into physiological, behavioural and biological characteristics (Yang & Nanni, 2011).

Also in **Figure 2**, the factors can be further subdivided into *biological biometrics* –representing the *inner part* of living beings –, *physiological biometrics* –morphological human factors–, and *behavioural biometrics,* referring to pattern of human behaviour (Jain, Ross and Pankanti, 2004). Thereby biological biometrics are, besides others, muscle contractions by deriving a Mechanomyogram (MMG) (Enamamu et al., 2017) , eye movement by means of Electrooculography (EOG), volumetric changes of blood in peripheral using Photoplethysmography (PPG) (Vhaduri & Poellabauer, 2017), bio-electrical signals such as cerebral waves of an Electroencephalogram (EEG), or electric muscles activities of the heart via an Electrocardiogram (ECG) (Singh & Singh, 2012).

Physiological biometrics are morphological factors (Idrus et al., 2013) such as the characteristics of the voice, the pattern of the iris, the geometry of the hand and fingerprints. In contrast, gestures, gait, keystroke dynamics, lip motion, pen pressure, signature geometry and speed are attributes related to the (learned) behaviour - congenital, trained and culturally acquired-, and are ascribed to *behavioural biometrics* (Bolle et al., 2004).

As the classification already indicates, there are many different biometrics that are also suitable for authentication, this is also shown by numerous current research studies on this topic: *Neal and Woodard* provide a survey for mobile device security in which they compare several physical biometrics such as fingerprint, iris, face and periocular as well as factors such as gait, keystroke dynamics, voice, and touch gestures (Neal & Woodard, 2016). Besides, *Mahfouz et al.* describe behavioural biometric authentication based on smartphones by measuring the gait with smartphone integrated sensors (Mahfouz, Mahmoud and Eldin, 2017). A similar approach is taken by *Lipps, Herbst and Schotten* as they build a system to analyse the gait with a sensor matrix located in the insole of a shoe (Lipps, Herbst and Schotten, 2021b). Furthermore, *Ruin and Yan* are highlighting potential risks, common network attacks and faking sensors by resubmitting biometric features. Their overview about existing biometric authentication systems include a rating system with respect to accuracy, efficiency, security and privacy features (Rui & Yan, 2018). With respect to MFA scenarios, *Hammad et al. p*ropose the

combination of ECG and fingerprint data, to obtain a multimodal biometric system (Hammad, Liu and Wang, 2018).

However, this line-up also indicates, that "biometrics are unique identifiers, but they are not secrets" (Schneier, 1999). People leave their biometric features such as fingerprints in many places, and some, such as the iris, facial characteristics or body shape, are easy to recognize even from the outside. Moreover, the loss of a factor, whether by theft (digital representation, validation feature) or by destruction (loss of a finger, change of friction grooves), implies the destruction and unusability of the feature forever.

But nevertheless, cryptographically, these biometric features are generally stronger than standard passwords; for example, a fingerprint contains significantly more information than a 128-bit passcode and given their always with you character -biometric features are always directly linked to the person to be authenticated- it is almost impossible to lose or forget such a biometric feature. The challenge in using it is twofold: i) validating that a factor belongs to the right person, and ii) that the factor matches the stored factor (Schneier, 1999), but this is where modern approaches such as Artificial Intelligence can assist.

3. Artificial Intelligence and its Applications

Alan Turing provided the so called Turing Test back in 1950 to provide a "[…] satisfactory operational definition of intelligence" (Russell & Norvig, 2016). Till this time there is no known complete AI which is able to satisfy all aspects of this test. In today's world in nearly every aspect of the modern life AI-techniques have at least some kind of approach. Be it in the fields of online shopping, advertising, enhanced Web searches, or in the topics of transportation, efficient manufacturing, or in the fields of food and farming. There is a huge variety of AI-subfields, ranging from general learning and recognising algorithms to specific ones for playing chess, proving mathematical theorems, writing poetry, driving cars, or for diagnostics of different diseases (Russell & Norvig, 2016). AI techniques encompass intelligent algorithms capable of performing certain tasks, sometimes better than a normal human would.

AI is used to create useful systems which are able to decide "intelligent" on the basis of the given data (Alpaydin, 2016). Therefor we use the continuous mathematics developed and discovered through the last centuries. Already in 1936 Turing showed that every possible computation can in principle be performed with a mathematical system. It was called the universal Turing machine (Russell & Norvig, 2016). At 1985 the thinking went one step further and it was said that "thinking and computing are radically the same thing" (Haugeland, 1985). Today state-of-the-art image or handwrite-recognise-systems are nearly as good at recognising specific objects or numbers as a normal human could do, in a much faster amount of time (Dodge & Karam, 2017).

According to *Margaret Boden* (Boden, 2016) there are five major types of AI, namely: Classical/Symbolic; Artificial Neural Networks; Evolutionary Programming; Cellular Automata and Dynamical Systems. The *Classical/Symbolic AI*, especially when combined with statistics, has its approaches in model learning, pattern detection, planning and reasoning. *Artificial Neural Networks* are especially good at pattern-recognition. The power of this system is that it finds necessary features by it´s own, although data is absolutely required. This data used will develop a system on which a trained algorithm will base its decisions. How wrong this can go showed the known amazon AI recruiting tool (Lauret, 2019) which preferred the male gender for recruiting, first without a clear reason. *Evolutionary Programming* is used to throw light on the biological evolution and brain development whereas *Cellular Automata* describes the field of discrete mathematical systems to characterize local interactions and dynamical computational evolutions. *Dynamical Systems* on the other hand can be used to model the development in living organisms.

For the approach of this paper the classical/ Symbolic AI was used to detect certain patterns. The Classical ML has three broad types: supervised; unsupervised and reinforcement learning (Boden, 2016). For supervised learning the system knows the patterns it has to predict. The system can be trained directly on the data. A system is then able to recognise a specific pattern, with the right selection of features at the training stage. Different mathematical algorithms and theories explored in the last centuries can be used for this approach. Unsupervised learning on the other hand means the system does not have a certain output, but tries to cluster and to find patterns with the given Data and features. Reinforcement learning means that a system gets trained with the help of rewards on certain training stages.

In approach of this work the system is trained with the help of supervised learning. Therefor the classification algorithms: K-Nearest Neighbor (KNN), Support Vector Machines (SVM) and Gaussian Naïve Bayes (GNB) are used and compared. The KNN classifier algorithm determines the distance between different points in a diagram. The closest datapoint to other points gets then classified as an associated class. The Letter "K" indicates how many neighbors are used for determining the distance. The SVM classifier algorithm uses supporting boundaries -vectors- between the point of certain classes -the hyperplanes-. The algorithm then tries to maximize the margin between the hyperplane and the supporting vectors which are at the position of certain data points. The GNB classifier algorithm is based on the Bayesian statistics and therefore with condition probabilities for not independent events. The GNB is the extension estimating the mean and standard deviation for certain training data (Berrar, 2019).

4. Measuring ECG Signals: The Experimental Setup

In terms of biometrical authentication, the ECG can be classified as a biological factor, since the signal is a visualization of the electrical activity of the heart. An advantage over physiological features is its robustness against any physiological changes. Simple everyday accidents and procedures such as cuts on the fingers and swelling of the face due to medical treatments would normally be problematic for fingerprint scanning methods or face recognition. With the rapid evolution of modern technology, accessibility of heart rate supervision is made possible with wearables such as smartwatches. Due to the characteristics of the ECG signal, a continuous authentication, desirable for the priory mentioned environment is possible.

The signal itself is a recording of the electrical activity of the heart, measured on the skin. It is observable with every heartbeat, consisting of the mechanical contraction of one heart chamber and the following relaxation. The resulting signal of one heartbeat is called a PQRST Wave, as depicted in **Figure 4**. The measurement can be done using either one or up to twelve electrodes, which need to be placed on specific points of the body. To reduce the complexity of the setup and improve the applicability in real life, a three-lead ECG is used for the experiment, which is a common setup for long-term measurements. Between each of the electrodes, the voltage is measured resulting in a vector according to the bipolar Einthoven-derivations.

For the recording and processing of data an Arduino Mega ADK Micro Control Unit (MCU) is connected to the ECG Click Sensor from MikroElektronika. The measured data is subsequently transmitted to the Arduino, which forwards it to the computer, where further processing and calculations are conducted using the programming language Python. Regarding a proof-of-concept data of 25 subjects are recorded, out of which 12 are male and 13 are female. In order to increase the integrity of the results similar conditions for the measurements of the individuals had to be ensured which is why the following aspects were considered. A challenging aspect regarding the recording process is the sensitivity of the cables and sensors. To ensure usable recordings, the data can only be acquired without the person moving. Furthermore, other environmental factors often cause noise. Those are for the instance the charging of the computer during the data acquirement. To avoid this, the subjects were measured twice for a duration of 5 minutes without prior exercise to achieve a resting heart rate during the measurement. Additionally, movement is kept to a minimum. For hygiene and reliability purposes, new pads were used for each individual. Further processing is done on a desktop PC following the steps shown in **Figure 3**.

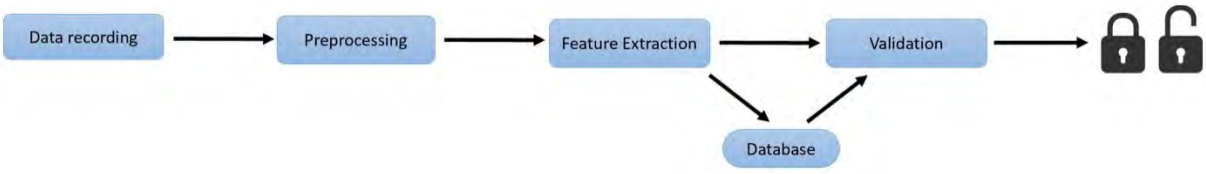

Figure 3: Steps of processing of the ECG data. After the data has been recorded, it is subjected to appropriate pre-processing and feature extraction.

5. Classification and discussion of the results

To validate the identity of an individual based on the ECG recordings, a classification of the data is necessary. In order to apply ML approaches, the recordings are pre-processed. In a first step, a filter is applied to reduce noise caused by the environment or movement of the person. Subsequently the recordings are segmented by finding characteristic waves, the so called PQRST Waves. One of these waves including its characteristic features are

depicted in **Figure 4**. Each segment includes 6 waves in order to counteract outliers and anomalies in the filtered data.

To keep the required storage space to a minimum and enable a comparison, features are extracted from each segment and stored in a feature vector. The applied fiducial feature extraction focusses on significant points in one PQRST wave. The resulting feature vector for one segment includes the medium values of the PQRST wave duration, R peak amplitude, T peak amplitude, QRS wave duration, P peak amplitude, Q peak amplitude, S peak amplitude and medium value of all data. These vectors are used as an input for the classification algorithms.

Comparing the measured data from 2 different subjects, significant differences are observable. **Table 1** includes two vectors of the same person A, extracted from distinct measurements and one feature vector from a different subject B. It is observable, that the duration of the QRS Wave does not differ significantly between the person A and B, while the amplitudes of the S Peaks are recognizably different. However, divergences also occur within the results of one subject. This becomes obvious looking at the different durations of the PQRST Wave in the given feature vectors from subject A. This observation is confirmed by **Figure 5** and **6**, with the left one showing two recordings of subject A, while the right one depicts one ECG signal of Subject A and one of B.

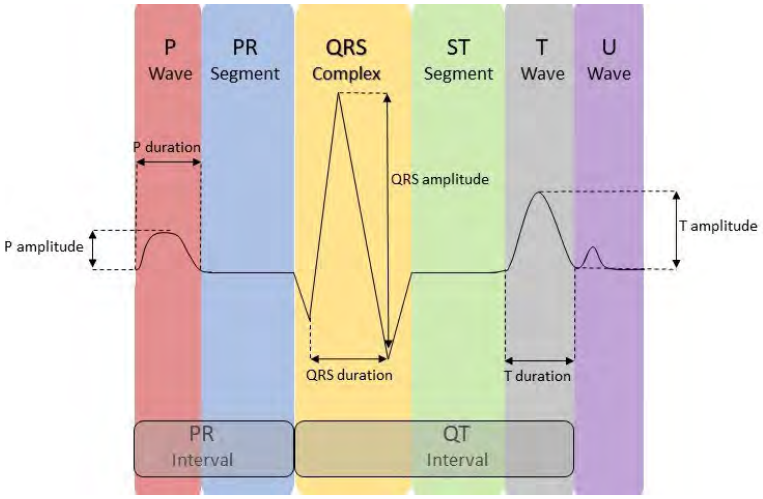

Figure 4: Characteristics of a PQRST Waves

Table 1: Feature Vectors of Recordings of different Subjects

Subject	PQRST Wave Duration	R Peak Duration	T Wave Amplitude	QRS Wave Duration	P Peak Amplitude	Q Peak Amplitude	S Peak Amplitude	Medium Value of all Data
A	661.44	639.66	286.18	102.92	220.14	163.04	79.80	196.86
A	625.70	641.34	279.39	102.92	217.46	162.54	77.79	196.31
B	730.67	448.38	236.45	102.19	214.06	190.04	153.55	205.28

Another challenge for the comparison are significant differences in the structure of the PQRST Wave of each individual person. They can differ in the number of occurring waves, as the S or P Wave is not always included, the orientation and order of the amplitude compared to the other waves.

Due to this drawbacks, three different ML approaches, namely the KNN Classifier, the GNB Classifier and the SVM, are applied, as they provide a higher level of adjustment to the data. Table 2 shows a comparison of the accuracies of these classification methods, depending on the database size.

Table 2: Comparison of the Classifiers

Number of Subjects	KNN Classifier	SVM	GNB Classifier
1	100%	100%	100%
2	100%	100%	100%
5	76.7%	97%	77%
10	62.3%	94%	61%
15	72%	94%	77%
20	37%	92%	90%

Number of Subjects	KNN Classifier	SVM	GNB Classifier
25	47%	94%	89%

The first approach is the KNN algorithm, which assigns the vectors to certain classes based on their distance to the clusters' centroids. This approach works for smaller sets, as for instance for two different persons a clear separation between the classes can be observed. However, problems occur on a larger database, which is caused by the higher rate of overlapping classes. This higher number of misclassified vectors can among others be reduced to the underlying distance measure, which is not adjusted to the specific properties of the data.

In comparison, the SVM is applied for the given set of data. The results of the SVM are more accurate than the ones for the k means algorithm, since the hyperplane, separating the different classes from each other, is more adjustable to the data. In the opposite, the distance measurement used for the KNN algorithm cannot respect outliers in the same degree. However, outliers and anomalies are also a challenge for the SVM, as the hyperplane might not be able to adjust all vectors to the right classes. However, the accuracies for the different database sizes show, that only few vectors are misclassified. In addition, the wrongly classified sample are recordings from different persons, where only one out of them is wrongly classified. Therefore, the possibility of wrongly accepting or rejecting a user in a real application is rather small. Although the percentage of the correct classified persons get less with a greater database, it is for a size up to 25 persons constantly over 90%. The third tested classification method is the GNB Classifier. Compared to the results for 25 individuals, it can be observed, that the number of misclassified points decreases with more people. Thus, the overall accuracy increases. Furthermore, only isolated outliers can be observed for 25 subjects, whereas results from 10 people show more misclassified data for the same person. In addition, the results for 10 persons show one subjects, where nearly all samples are classified wrongly. This is an unsatisfying result, since it would cause a DoS for person in an established authentication system or would grant access to the wrong person.

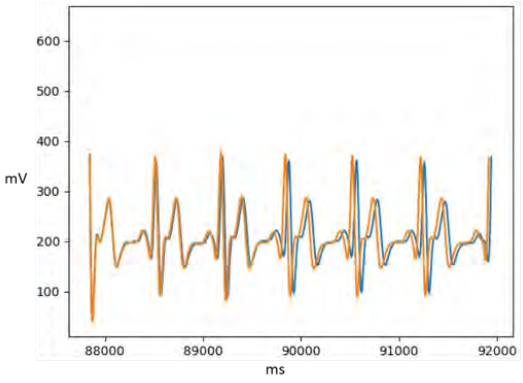

Figure 5: ECG Signals of one Individual

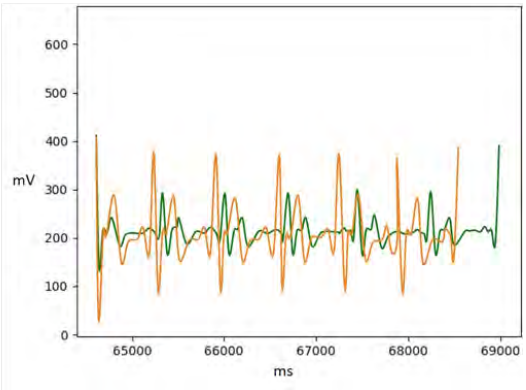

Figure 6: ECG Signals of two Individuals

Considering the feature vectors of these two subjects, it is revealed that all entries of the vectors only slightly differ from the other person's ECG signal. This shows, that although the GNB considers each feature individually, it gets problems with small datasets and similar samples of two subjects. However, for a larger set of data, smaller differences in the vectors gain importance. Hence, the GNB classifier provides a more accurate result for the data set including data from 25 people.

Compared to the KNN algorithm, the GNB Classifier achieves rather similar accuracies for smaller sets of data, while for a larger database, it outperforms the clustering algorithm significantly. The SVM achieves the best and most stable results for all database sizes given in the table. However, for a set of 25 people, the GNB classifier nearly achieves the same accuracy. This observation can be reduced to the underlying algorithms. As mentioned before, the Euclidean Distance is used in the KNN algorithm as a measurement for the differences, whereas the measurement of the other two approaches are adjusted to the data. Thus, its performance is less satisfying. In addition, a disadvantage of the KNN algorithm are the problems with the classification of non-globular shaped data clusters. Compared to this, the GNB classifier considers the Gaussian Distribution for each individual feature, comparing all values of the vector separately from each other. However, it assumes, that all features are independent of each other, which cannot be satisfied, as the duration of the overall wave and the smaller waves it includes are obviously depending. The SVM puts the focus on the interaction of the values instead of comparing them separately from each other. Therefore, it outperforms the GNB Classifier when comparing two

samples with similar values at specific entries in the feature vector, as a combination of all values is used for the comparison. Thus, the SVM results in the highest accuracy, considering the given database including the samples of 25 persons and the chosen separation of trainings and test data. However, these conclusions are made considering a rather small data set. To determine the accuracies of a larger database, tests have yet to be conducted.

6. Outlook and Future Work

The use of biometric factors for the authentication of human entities is shifting back into focus due to current possibilities and modern sensor technology. Bygones are the days of simple fingerprints or iris scanners, especially since many of these approaches are no longer secure. Methods such as gait recognition, behaviour-based identification and the utilization of Electrocardiogram values are of growing relevance.

For this purpose, this work describes a setup and proof-of-concept for ECG-based authentication. With a simple test-set-up based on Arduino Microcontroller, ECG click shield and ECG-pads the values of 25 persons are recorded, processed and evaluated. Three different Machine Learning methods -K-Nearest Neighbor, Support Vector Machines and Gaussian Naïve Bayes-, were used and compared with respect to their suitability and performance. The results show that for only a few individuals all algorithms perform equally well, which was to be expected. As the number of participants increases, only the SVM delivers very good results and the GNB classifier still delivers good results. After a dent in performance, however, the KNNs are also improving again. The work thereby demonstrates that ECG-based authentication works in principle and can be used as a factor for authentication.

Due to the simple test setup, only ECG values with persons at rest could be recorded. In the next step, with improved hardware, persons will also be examined during other activities such as walking or light and medium physical exertion. It is expected that the heart values are even more significant. Improved hardware also includes the use of a 1-lead ECG, instead of the 3-lead ECG used in this case, for example with a sensor on the wrist.

Because the ML methods used performed differently, it is necessary to investigate whether a combination of the methods might provide better results. In addition, other ML methodologies will also be included from the areas supervised-, unsupervised, and reinforcement learning.

Acknowledgment

This work has been supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (Förderkennzeichen 16KISK003K Open 6G Hub). The authors alone are responsible for the content of the paper.

References

Alpaydin, E., "Machine Learning: The New AI"*, The MIT Press Essential Knowledge Series*, ISBN: 978-0262529518, 2016.

Barricielli, B.R., Casiraghi, E., and Fogli, D., "A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications", *IEEE Access,* Issue 7, pp. 167653 -- 167671, DOI: 10.1109/ACCWSS.2019.2953499, 2019.

Berrar, D.P., "Bayes' Theorem and Naive Bayes Classifier", *Encyclopedia of Bioinformatics and Computational Biology*, DOI: 10.1016/B978-0-12-809633-8-20473-1, 2019.

Boden, M.A., "AI - Its nature and future", *Oxford University Press*, ISBN: 978-0198777981, 2016.

Bolle, R.M., Connell, J.H., Pankanti, S., Ratha, N.K., and Senior, A.W., "Guide to Biometrics", Springer Professional Computing, Springer, DOI: 10.1007/978-1-4757-40036-3, 2004.

Burr, W.E., Dodson, D.F., Nweton, E.M., Perlner, R.A., Polk, W.T., Gupta S., and Nabbus, E.A., "Electronic Authentication Guidline - NIST Special Publication 800-63-2", National Institute of Standards and Technology, DOI: 10.6028/NIST.SP.800-63-2, 2017.

Dodge, S., and Karam, L., "A Study and Comparison of Human and Deep Learning Recognition Performance Under Visual Distortions"*, Cornell University - Computer Science - Computer Vision and Pattern Recognition,* arXiv:1705.02498, 2017.

Duque Anton, S., Fraunholz, D., Lipps, C., Alam, K, and Schotten, H.D., "Putting Things in Context: Securing Industrial Authentication with Context Information", *International Journal on Cyber Situational Awareness (IJCSA), vol.* 3, no. 1, DOI:10.226119/IJCSA.2018.100122, 2018.

Enamamu, T.S., Clarke, N., Haskell-Dowland, P., and Li, F., "Transparent authentication: Utilising heart rate for user authentication", *12th International Conference for Internet Technology and Secured Transactions (ICTST),* Cambrige, United Kingdom, DOI: 10.23919/ICTST.2017.8356401, 2017

Hammad, M., Liu, Y., and Wang, K., "Multimodal Biometric Authentication Systems Using Convolution Neural Network Based on Different Level Fusion of ECG and Fingerprint", *IEEE Access,* Issue 7, pp. 26527-26542, DOI: 10.1109/ACCESS.2018.2886573, 2018

Haugeland, J., "Artificial Intelligence: The Very Idea", *The Massachusetts Institute of Technology,* ISBN: 978-0262081535, 1985.

Idrus, S.Z.S., Cherrier, E., Rosenberger, C., and Schwartzmann, J.-J., "A Review on Authentication Methods", *Australian Journal of Basic and Applied Sciences,* 7(5), pp. 95-107.

Jain, A., Ross, A., and Pankanti, S., "Biometrics: a tool for information security", *IEEE Transactions on Information Forensics and Security ,* vol. 1, no. 2, pp. 125-143., DOI: 10.1109/TIFS.2006.873653

Jain, A., Ross, A., and Prabhakar, S., "An Introduction to Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology ,* vol. 14, no. 1, pp. 4-20, DOI:10.1109/TCSVT.2003.818349, 2004

Lauret, J., "Amazon's sexist AI recruiting tool: how did it go so wrong?", [Online] Available at: https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e, 2019, [Zugriff am 30 10 2021].

Lipps, C., Ahr, P., and Schotten, H. D., "The PhySec Thing: About Trust and Security in Industrial IoT Systems", *Journal of Information Warfare* , vol. 19, no. 3, pp. 35-49, 2020a

Lipps, C., Bergkemper, L., and Schotten, H. D., "Distinguishing Hearts: How Machine Learning identifies People based on their Heartbeat", *6th International Conference on Advances in Biomedical Engineering (ICABME2021),* DOI:10.1109/ICABME53305.2021.9604855, 2021a

Lipps, C., Herbst, J., and Schotten, H. D., "How to Dance Your Passwords: A Biometric MFA-Scheme for Identification and Authentication of Individuals in IIoT Environments", 16th International Conference on Cyber Warfare and Security (ICCWS), Cookeville, TN, USA, 2021b.

Lipps, C., Mallikarjun, S.B., Strufe, M., Heinz, C., Grimm, C., and Schotten, H.D., "Keep Private Networks Private: Secure Channel-PUFs, and Physical Layer Security by Linear Regression Enhanced Channel Profiles", 3rd International Conference on Data Intelligence and Security (ICDIS), South Padre Island, TX, USA, DOI: 10.1109/ICDIS50059.2020.00019, 2020b

Lipps, C., Weinand, A., Krummacker, D., Fischer, C., and Schotten, H.D., "Proof of Concept for IoT Device Authentication Based on SRAM PUFs Using ATMEGA 2560-MCU", *1st International Conference on Data Intelligence and Security*, South Padre Island, TX, USA, 2018, DOI: 10.1109/ICDIS.2018.00013

Mahfouz, A., Mahmoud, T.M., and Eldin, A.S., "A survey on behavioral biometric authentication on smartphones", *Journal of Information Security and Applications,* vol. 37, pp. 28--37, 2017, DOI:10.1016/j.jisa.2017.10.002.

Neal, T.J., and Woodard, D.L., "Surveying Biometric Authentication for Mobile Device Security", *Journal of Pattern Recognition Research,* vol. 11, no. 1, pp. 74--110, 2016, DOI: 10.13176/11.764.

O'Gorman, L., "Comparing passwords, tokens, and biometrics for user authentication", *Proceedings of the IEEE, vol.* 91, no. 12, pp. 2021--2040, DOI: 10.1109/JPROC.2003.819611.

Ometov, A., Bezzateev, S., Mäkitalo, N., Andreev, S., Mikkonen, T., and Koucheryavy, Y., "Multi-Factor Authentication: A Survey", *cryptography,* MDPI, vol. 1, no. 2, 2017, DOI: 10.3390/cryptography2010001.

Renaud, K., "Evaluating Authentication Mechanisms" *Security and Usability: Designing Secure Systems That People Can Use,* O'Reilly, pp. 103-128.

Rui, Z., and Yan, Z., "A Survey on Biometric Authentication: Toward Secure and Privacy-Preserving Identification", *IEEE Access,* Issue 7, pp. 5994-6009, DOI:10.1109/ACCESS.2018.2889996.

Russell, S.J., and Norvig, P., "Artificial Intelligence: A Modern Approach*", Pearson,* 3rd Edition, Addison Wesley, 2009, ISBN: 978-0136042597

Schäfer, G., and Rossberg, M., "Security in Fixed and Wireless Networks*",* 2nd Edition, Heidelberg, Germany, John Wiley & Sons Ltd., 2016, ISBN: 978-1119040743.

Schneier, B., "The Uses and Abuses of Biometrics" *Communications of th ACM,* vol. 42, no. 8, p. 136.

Schneier, B., "Carry on: Sound Advice from Schneier on Security"*,* 1st Edition, Indianapolis, Indiana, United States, John Wiley & Sons Inc., 2014, ISBN: 978-1118790816.

Schneier, B., "Applied Cryptography: Protocols, Algorithms and Source Code in C*",* 20th Anniversary Edition, Indianapolis, IN, USA, John Wiley & Sons Ltd., 2015, ISBN: 978-1-119-09672-6.

Singh, Y.N., and Singh, S., "Evaluation of Electrocardiogram for Biometric Authentication", *Journal of Information Security,* vol. 3, no. 1, 2012, DOI:10.4236/jis.2012.31005.

Vhaduri, S., and Poellabauer, C., "Wearable device user authentication using physiological and behavioral metrics*",* IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 2017, 10.1109/PIMRC.2017.8292272.

Yang, J., and Nanni, L., "State of the art in Biometrics*",* 1st Edition, London, United Kindom, IntechOpen Limited, 2011, ISBN: 9789535155669.

Zhou, B., Singh, M.S., Doda, S., Yildrim, M., Chengm J., and Lukowicz, P., "The Carpet Knows: Identifying People in a Smart Environment from a Signle Step", I*EEE International Conference on Pervasive Computing and Communications (PerCon-17) - First International Workshop on Pervasive Smart Living Spaces*, Kona, Hawaii, United States, 2017, 10.1109/PERCOMW.2017.7917618.